Statistical Analysis of Disease Onset and Lifetime Data from Tumorigenicity Experiments

By S. W. Lagakos* and Louise M. Ryan*

We present and discuss several methods for analyzing rodent tumorigenicity experiments. Two approaches are based on the age and tumor status (present/absent) of each animal at the time of death, and assume either that the tumor type is nonlethal or instantly lethal. Two other approaches avoid such restrictive assumptions about tumor lethality by requiring additional types of data. One method assumes that animals are randomly sacrificed at various ages throughout the study. The second approach requires that each animal which develops the tumor be classified as dying either from the tumor or from other causes.

Introduction

Rodent tumorigenicity experiments are frequently used to assess the safety of chemicals, food additives, pesticides, and other products. Typically, a control group of rodents is compared with one or more "exposed" groups that are fed the compound of interest throughout their lifetimes. The substance is deemed carcinogenic if, loosely speaking, it increases the rate of development of one or more tumor types.

From a statistical point of view, the analysis of rodent tumorigenicity experiments poses a number of interesting and challenging problems. One reason for this is that most tumor types are occult and therefore detectable only after the animal has died. The statistical analysis of experiments involving observable or palpable tumors, such as skin and mammary tumors, is much simpler and will not be discussed in this paper. Thus, the quantity of interest, time to tumor onset, is not directly observed—we know only whether or not the tumor has developed by the time of death. The problem is compounded by two other phenomena: first, tumor types differ in their lethality, with some causing death shortly after onset and others never or rarely causing death. This means that the time from onset to death will depend in part on the tumor lethality. The second complication is that because of the extremely high dosages of the test compounds, the longevity of the exposed animals is often affected. Consequently, tumors might appear to occur earlier in the exposed group simply because death occurs sooner, or might not appear at all because the animals died before the tumor could develop.

Because of these complexities, the comparison of control and exposed groups based on the proportions of animals that eventually develop tumors can be misleading. Although it is now widely recognized that comparisons between the control and exposed groups should account for age at death, there is no single appropriate method of incorporating age. We shall see that appropriate methods exist when the tumor type in question is nonlethal, when it is instantly lethal, when there are serial sacrifices, or when pathologists determine whether or not the detected tumors caused death.

This paper reviews these methods and discusses their advantages and shortcomings.

General Formulation

Figure 1 gives a simple representation of the life of an animal in a tumorigenicity experiment. Animals be-
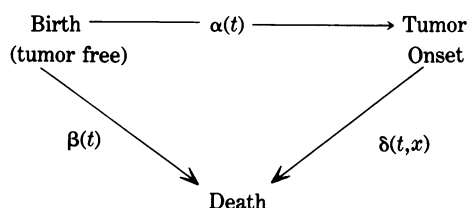

FIGURE 1. General model.

*Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115.

gin in a tumor-free state and then either develop the tumor or die tumor-free. By "tumor onset" we mean the earliest age at which the tumor would be detectable by a microscopic examination of the involved organ or tissue. Animals that develop the tumor eventually die with the tumor present.

Suppose U_1 denotes the age at (time to) the first of tumor onset or death, $a(t)$ indicates whether [$a(t) = 1$] or not [$a(t) = 0$] the tumor has occurred by age t, and U denotes age at (time to) death. Probabilistically, the model can be characterized by the intensity functions $\alpha(t)$, $\beta(t)$, and $\delta(t,x)$, where

$$\alpha(t) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \mathrm{Pr}\{t \le U_1 < t + \Delta t,$$
$$a(U_1) = 1 \mid U_1 \ge t\}$$

$$\beta(t) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \mathrm{Pr}\{t \le U_1 < t + \Delta t,$$
$$a(U_1) = 0 \mid U_1 \ge t\}$$

and

$$\delta(t,x) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \mathrm{Pr}\{t \le U < t + \Delta t \mid U \ge t,$$
$$U_1 = x\}$$

The first of these functions, $\alpha(t)$, specifies the risk of tumor onset and is the quantity of primary interest in the comparison of the control and exposed groups. The function $\beta(t)$ describes the risk of death without tumor. In many experiments, this risk is affected by the high doses of compound that are given to the rodents; consequently, differences between the control and exposed groups with respect to the function $\beta(t)$ are not regarded as evidence of tumorigenicity. The function $\delta(t,x)$, which describes the death risk with tumor, is also a nuisance function, because it also may depend upon the toxicity of the chemical being tested. The statistical problem is to test the equality of the functions $\alpha(t)$ in the control and exposed groups in the presence of the nuisance functions $\beta(t)$ and $\delta(t,x)$.

A function which is often considered to be a measure of carcinogenicity is the tumor prevalence rate, or the probability that an animal living at age t has developed the tumor; that is,

$$\pi(t) = \mathrm{Pr}\{a(t) = 1 \mid U \ge t\} \tag{1}$$

In terms of Figure 1, the prevalence function can be expressed as

$$\pi(t) = \int_0^t g(t,x)dx/\{1 + \int_0^t g(t,x)dx\} \tag{2}$$

where

$$g(t,x) = \alpha(x) \exp\left\{\int_x^t [\alpha(u) + \beta(u) - \delta(u,x)]du\right\}$$

The observable data for each animal consist of U, the age at death, and the binary variable $a = a(U)$, which indicates whether ($a = 1$) or not ($a = 0$) the tumor has developed by the time of death. The only estimable aspects of the model depicted in Figure 1 are those functions that can be expressed in terms of the distribution of the observed random variables (U,a). The latter are characterized (*1*) by the cause-specific hazard functions

$$h_1(t) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \mathrm{Pr}\{t \le U < t + \Delta t, a = 1 \mid U \ge t\}$$

and

$$h_0(t) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \mathrm{Pr}\{t \le U < t + \Delta t, a = 0 \mid U \ge t\}$$

It can be verified that

$$h_1(t) = \int_0^t g(t,x) \, \delta(t,x)dx/\{1 + \int_0^t g(t,x)dx\}$$

and

$$h_0(t) = \beta(t)/\{1 + \int_0^t g(t,x)dx\}$$

The fundamental difficulty in the analysis of carcinogenicity experiments is that none of $\alpha(t)$, $\beta(t)$, or $\delta(t,x)$ can be expressed solely in terms of $h_1(t)$ or $h_0(t)$. Thus, without further assumptions, these three basic model parameters are nonidentifiable when the data consist only of U and a. Furthermore, since $h_0(t)$ and $h_1(t)$ depend on the nuisance functions $\beta(t)$ and $\delta(t,x)$, neither is appropriate as a basis for determining tumorigenicity. The rest of the paper will discuss proposed methods for overcoming this nonidentifiability problem.

Two Special Cases: Instantly Lethal and Nonlethal Tumor Types

Instantly Lethal Tumor Types

For two special cases of the general model, appropriate analyses of $\alpha(t)$ are possible. The first of these is when the tumor type under consideration is instantly lethal; i.e., death follows instantaneously after tumor onset. For example, leukemias and reticulum cell sarcomas are considered to be virtually instantly lethal. For such tumors, the process in Figure 1 is characterized solely by the functions $\alpha(t)$ and $\beta(t)$, and simplifies to one where an animal begins tumor free and then either develops a tumor (and dies instantly), or dies without a tumor. The standard competing-risks frame-

work can be used to describe this process by taking

$$U = \min(T_1, Y)$$

and

$$a = \begin{cases} 1 & \text{if } T_1 \leq Y \\ 0 & \text{if } T_1 > Y \end{cases}$$

where T_1 and Y are random variables representing the potential times to tumor onset and death without tumor, respectively. These are imagined to "compete" with one another to determine when and how death occurs.

Analyses of competing-risk data are well-documented in the statistical literature (2). For example, the cumulative intensity function

$$A(t) = \int \alpha(u)du$$

can be consistently estimated by

$$\Sigma \, n(t)/N(t)$$

where $n(t)$ is the number of tumor deaths at age t and $N(t)$ is the number of animals alive at age $t - 0$.

Perhaps the most common test of equality of groups with respect to the functions $\alpha(t)$ is the logrank test. This test arises as the partial likelihood score test of $\theta = 0$ in the proportional hazards model where

$$\alpha_E(t) = \exp(\theta) \, \alpha_C(t)$$

and the subscripts denote the exposed and control groups, respectively. The construction of the test is described in Figure 2. Essentially, the control and exposed groups are compared at each age of death where there is a tumor, and the observed and expected number of tumor deaths in the exposed group are summed over these death times. The test is valid regardless of whether or not $\alpha_E(t)$ is proportional to $\alpha_C(t)$, but is most efficient when proportionality holds. Other so-called "lifetable" tests are designed to be most sensitive to alternatives other than proportionality. When the tumor type is not instantly lethal, the noncentrality parameter of the logrank and related tests are distorted in a direction that depends on the functions $\beta_E(t)$ and $\beta_C(t)$ (3). In particular, when the risk of death from nontumor causes is greater in the exposed group, the resulting significance levels tend to be too small. That is, when applied to tumors which are not instantly lethal, the logrank test is invalid and tends to reject the null hypothesis too frequently.

Nonlethal Tumor Types

Formally, a nonlethal tumor type is one for which $\delta(t,x) = \beta(t)$ for all t and x; that is, a tumor type for

Data at Age t

	Dead with tumor		Number at risk
Controls	m	$M - m$	M
Exposed	n	$N - n$	N
Total	$(m + n)$		$(M + N)$

FIGURE 2. Construction of logrank and Hoel-Walburg tests. *Note*: for logrank test, M and N denote number of animals that survive until at least age t. For Hoel-Walburg test, M and N denote number of animals that die at age t. Both test statistics are of the form: $T = \Sigma(O_t - E_t)/(\Sigma V_t)$, where $O_t = n$, $E_t = N(m + n)/(M + N)$, and $V_t = MN \, (m + n)(M + N - m - n)/[(M + N)^2 \, (M + N - 1)]$.

which tumor onset does not alter the risk of death. The identifiable functions $h_1(t)$ and $h_0(t)$ simplify to

$$h_1(t) = \beta(t) \, \pi(t)$$

and

$$h_0(t) = \beta(t) \{ 1 - \pi(t) \}$$

and the prevalence function becomes

$$\pi(t) = 1 - \exp\{-\int \alpha(u)du\}$$

Thus, both $\alpha(t)$ and $\beta(t)$ (and π) are identifiable.

Nonlethal tumor types can also be represented by a special competing risks framework, with

$$U = Y$$

and

$$a = \begin{cases} 1 & \text{if } T_1 \leq Y \\ 0 & \text{if } T_1 > Y \end{cases}$$

where the independent random variables T_1 and Y represent the potential time to tumor onset and the time to death, respectively. Note that T_1 is either right censored (when $a = 0$) or left censored (when $a = 1$); it is never directly observed. Also, $\pi(t)$ can be thought of as the c.d.f. of T_1, with corresponding hazard function $\alpha(t)$. This representation shows that the problem of estimating $\pi(t)$ [or, equivalently $\alpha(t)$] corresponds to the so-called binomial extremum problem (4); i.e., let $t_1 < t_2 < \ldots$ denote the distinct ages of death, let N_j be the number of deaths at age t_j, and let n_j be the number of

these with a tumor. Then, conditional on N_1, N_2, ..., n_1, n_2, ... are independently binomial with probabilities $\pi(t_j)$ that are ordered. It can be shown (4,5) that the constrained maximum likelihood estimator of the set of π_j is also the solution to the constrained weighted least-squares criterion

$$\Sigma \, N \, \{\pi \, (t_j) - n_j/N_j\}$$

and given by the isotonic regression of the naive estimators n_j/N_j with weights N_j.

It can be verified that

$$\pi(t) = \Pr\{\, a(U) = 1 \mid U = t \,\} \qquad (3)$$

which indicates that animals dying at age t are representative (6), with respect to tumor presence, of all animals surviving until at least age t. It follows that a test for differences between $\pi_E(t)$ and $\pi_C(t)$ can be obtained by comparing the control and exposed groups with respect to the proportions of tumors found in animals dying at age t. The first such "prevalence" test was proposed by Hoel and Walburg (7) and has the same form as the logrank test, except that the numbers of animals "at risk" at age t are defined differently (see Fig. 2).

It can be shown (8) that the Hoel-Walburg test arises as the likelihood score test of $\theta = 0$ from the logistic model where

$$\text{logit}\{\pi_E(t)\} = \theta + \text{logit}\{\pi_C(t)\}$$

that is, where the two groups have proportional prevalence-odds functions. Thus, the Hoel-Walburg test is most efficient when the two prevalence functions have proportional odds, but remains valid when they are not proportional. Analogous to logrank tests, prevalence tests are distorted when applies to tumor types that are not nonlethal (6). Other types of prevalence tests have been recently investigated by McKnight and Crowley (9) and Finkelstein (10).

Use of Serial Sacrifices

For the special cases of the preceding section, problems of nonidentifiability of model parameters were avoided by assuming that the tumor type either is instantly lethal or that it is nonlethal. However, most types of tumors are somewhere between these two extremes. One approach to achieving identifiability for these intermediate tumor types utilizes serial sacrificing; that is, the random selection, killing, and examination of live animals at various ages. From Eq. (1), the proportion of animals sacrificed at age t that are found to have a tumor estimates $\pi(t)$, the tumor prevalence at age t. Thus, with serial sacrificing the identifiable aspects of the process depicted in Figure 1 are those expressible in terms of $h_1(t)$, $h_0(t)$, and $\pi(t)$. It
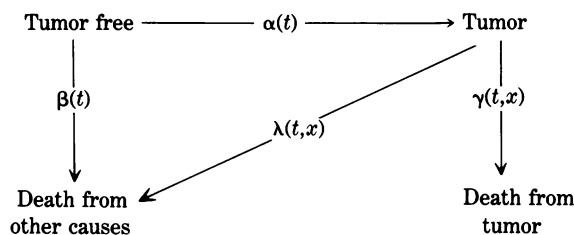

FIGURE 3.   Modified model incorporating cause of death.

follows that $\beta(t)$ is identifiable since we can write

$$\beta(t) = h_0(t)/[\,1 - \pi(t)\,]$$

It can also be shown that the function $\alpha(t)$ is identifiable, though a simple functional form such as that for $\beta(t)$ is not obtainable.

Although serial sacrificing can in theory overcome nonidentifiability problems, the methods (9,11–14) which use this information require very large amounts of data. For this reason, experiments with extensive sacrificing, other than at the termination of an experiment, are quite rare.

Use of Cause of Death Information

An alternative to using sacrifice data for estimating and testing time to tumor onset becomes available when each animal found to have a tumor at death is further classified as having died from that tumor or from non-tumor causes (see Figure 3). The observable "data" are (U,a,b), where the binary indicator variable b equals one for deaths due to the tumor and zero otherwise. This amounts to decomposing the intensity function $\delta(t,x)$ in Figure 1 into the sum $\gamma(t,x) + \lambda(t,x)$, where

$$\gamma(t,x) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \Pr\{t \leq U < t + \Delta t,$$
$$b = 1 \mid U \geq t, a = 1, U_1 = x\}$$

and

$$\lambda(t,x) = \lim_{\Delta t \to 0} (\Delta t)^{-1} \Pr\{t \leq U < t + \Delta t,$$
$$b = 0 \mid U \geq t, a = 1, U_1 = x\}$$

If a tumor develops at time x, $\gamma(t,x)$ and $\lambda(t,x)$ represent the risks of dying from the tumor and from other causes, respectively, at time t.

Suppose we assume that $\lambda(t,x) = \beta(t)$ for all t and x; that is, that the risk of death from nontumor causes is unaltered by tumor onset. Given this assumption and the data (U,a,b), the identifiable components of the model are the functions $\alpha(t)$, $\beta(t)$, and $\pi(t)$. The identifiability of $\pi(t)$ follows because it can be verified that under the above assumption

$$\pi(t) = \mathrm{Pr}\{\, a = 1 \mid U = t,\, b = 0 \,\}$$

Note also from Eq. (2) that the prevalence depends only on $\alpha(t)$ and $\gamma(t,x)$. Thus, the assumption that $\lambda(t,x) = \beta(t)$ essentially ensures that nontumor deaths act like sacrifices from the simpler model where the animal develops a tumor and at some later age dies. The only difference is that now the ages of death are not determined by the experimenter. Several methods for estimating these functions have been investigated (*15–18*).

A method of comparing the control and exposed groups for this model was proposed by Peto (*3,9,15,19*). The procedure consists essentially of applying of both the logrank and Hoel-Walburg tests to subsets of the data. Specifically, one first ignores all animals where death was due to the tumor, and applies a Hoel-Walburg test to the remainder. Next, one uses all the observations, treating deaths from the tumor as "uncensored" events, and all other deaths as "censored," and applies a logrank test. It can be shown that the former tests the equality of the control and exposed groups with respect to the tumor prevalence function $\pi(t)$, and the latter compares them with respect to the function

$$h^*(t) = \int_0^t g(t,x)\,\gamma(t,x)dx\Big/\Big\{1 + \int_0^t g(t,x)dx\Big\}$$

An overall test for differences between $\alpha(t)$ or $\gamma(t,x)$ in the control and exposed groups is obtained by combining the p-values from the two procedures.

Unlike the earlier tests that were discussed, Peto's approach requires no assumptions about tumor lethality, but instead requires accurate assignment of cause of death and assumes that $\beta(t) = \lambda(t,x)$. We carried out an empirical study (*6*) of the recently completed ED_{01} experiment and found these assumptions to be violated in various degrees for several tumor types. Some analytic results suggest however, that the effects of such violations on the testing problem will be less severe than the distortions caused to estimates of prevalence. We feel that a better understanding of the way in which pathologists define cause of death can lead to improvements in the analysis of carcinogenicity experiments.

Defining Cause of Death to Avoid Distortion

Death is an end result of many complex events. While it may sometimes be evident that the tumor type in question contributed to death, the definition of cause of death is necessarily somewhat arbitrary. Recall that any rule for assigning cause of death defines a particular model of the form in Figure 3, where the hazard δ is divided into two components:

$$\delta(t,x) = \lambda(t,x) + \gamma(t,x)$$

As discussed in the previous section, the two-step Peto method is valid when $\beta(t) = \lambda(t,x)$ and cause of death is accurately classified. Consideration of the observed violations of these assumptions in the ED_{01} data raises the possibility that statisticians work with pathologists to redefine the notion of "cause-of-death" (perhaps a different phrase ought to be used) so that the condition $\beta(t) = \lambda(t,x)$ holds, regardless of whether the resulting definition has complete biologic meaning in terms of actual causation. If this could be done, we would be assured that exposed and control groups could be accurately compared with respect to $\pi(t)$ and $h^*(t)$. It remains to be seen whether an effective modification of the conventional notion of cause of death can be found.

Another approach, which we have begun to explore, is to move away from the notion of cause of death altogether, and to incorporate information on the anatomic stage of the tumor. This approach has the potential to simplify the task of pathologists and to solve the nonidentifiability problem without the need for restrictive assumptions.

REFERENCES

1. Prentice, R. L., Kalbfleish, J. D., Peterson, A., Flourney, N., Farewell, V., and Breslow, N. The analysis of failure time in the presence of competing risks. Biometrics 34: 541–554 (1978).
2. Kalbfleish, J. D., and Prentice, R. L. The Statistical Analysis of Failure Time Data. Wiley, New York, 1980.
3. Lagakos, S. W. An evaluation of some two sample tests used to analyze animal carcinogenicity experiments. Util. Math. 21B: 239–260 (1982).
4. Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. Statistical Inference Under Order Restrictions. Wiley, New York, 1972.
5. Turnbull, B. W. The empirical distribution function with arbitrarily grouped, censored, and truncated data. J. Roy. Statist. Soc. B38: 290–295 (1976).
6. Lagakos, S. W., and Ryan, L. M. On the representativeness assumption in prevalence tests for carcinogenicity, J. Roy. Statist. Soc. C34: 54–62 (1984).
7. Hoel, D., and Walburg, H. E. Statistical analysis of survival experiments. J. Natl. Cancer Inst. 49: 361–372 (1972).
8. Dinse, G. E., and Lagakos, S. W. Regression analysis of tumor prevalence data. Applied statistics Vol 32, No 3: 236–248 (1983).
9. McKnight, B. and Crowley, J. Tests for differences in tumor incidence based on animal carcinogenicity experiments. J. Am. Statist. Assoc. 79, No. 387: 639–648 (1984).
10. Finkelstein, D. M. A proportional hazards model for arbitrarily interval censored survival data. Technical Report 388Z, Dana Farber Cancer Institute, Boston, MA 02115, 1984.
11. Dewanji, A., and Kalbfleish, J. D. Non-parametric methods for survival/sacrifice experiments. Paper presented at ASA joint meetings, Toronto, Canada, August 1983.
12. McKnight, B. Validity and efficiency in inferences about the tumor incidence rate. Paper presented at the ASA joint meetings, Philadelphia, PA, August 1984.
13. Anderson, K. Nonparametric estimation for survival/sacrifice experiments without assumptions about competing risks. Unpublished manuscript, Harvard School of Public Health, 1984.
14. Kodell, R. L. Utilizing cause-of-death information to estimate conditional probabilities of disease and death. Paper presented

at the ASA joint meetings, Philadelphia, PA, August 1984.

15. Peto, R., Pike, M. C., Day N. E., Gray, R., Lee, P., Parish, S., Peto, J., Richards, S., and Wahrendorf, J. Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In: IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemical to Humans, Supplement of Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal, IARC, Lyon, 1980, pp. 331–426.

16. Kodell, R. L., Shaw, G. W., and Johnson, A. M. Nonparametric joint estimators for disease resistance and survival functions in survival/sacrifice experiments. Biometrics 38: 43–58 (1982).

17. Dinse, G. E., and Lagakos, S. W. Nonparametric estimation of lifetime and disease onset distributions from incomplete data. Biometrics 38: 921–932 (1982).

18. Turnbull, B. W., and Mitchell, T. J. Nonparametric estimation of distributions of time to onset and time to death for specific diseases in survival/sacrifice experiments. Biometrics 40: 41–50 (1984).

19. Peto, R. Guidelines on the analysis of tumor rates and death rates in experimental animals. Brit. J. Cancer 29: 101–105 (1974).